Discrimination in the EU

Analysis Report

University Carlos III of Madrid

Master in Computational Social Science

Survey Research Methodology II

2023/2024

Clara Fernanda Martín Villaescusa

Laura Martínez Temiño

Ana Pérez Barrera

Table of contents

1. Introduction

In a speech on 25 September 2018, Antonio Guterres, Secretary-General of the United Nations, highlighted the need to intensify efforts to combat criminalisation, prejudice and violence based on sexual orientation, gender identity and sex characteristics (*United Nations Resolutions on Sexual Orientation, Gender Identity and Sex Characteristics*, n.d.). This statement underscores the importance of addressing human rights violations that directly affect the transgender community, a group that faces unique and significant challenges in society.

Discrimination against transgender people manifests itself directly and indirectly in the European Union (EU), affecting critical areas such as employment, education, health, and access to social services. While the acronym LGBTI serves to unify the community, it is crucial to recognise that transgender people represent a distinct group with specific needs, particularly regarding legal recognition of their gender identity.

The European Commission, recognising these challenges, is committed to promoting equality for transgender people. The Commission's 'Checklist of actions to promote LGBTI equality' of December 2015 is a clear example of policy efforts aimed at combating discrimination and promoting equality in multiple areas of social and economic life (‘Legal Gender Recognition in the EU’, 2020). Furthermore, the influence of European institutions and European political parties (Europarties) in promoting inclusive social and human rights policies is a vital component in the fight against discrimination (Van Der Vleuten, 2020). Through international socialisation policies, these entities can promote equality and non-discrimination norms that directly benefit the transgender community, showing how the integration of transnational policy agendas can advance gender equality and sexual diversity at the European level.

At the global level, the 2011 report of the Office of the High Commissioner for Human Rights and the 2015 joint statement of 12 UN bodies highlight the urgency of ensuring legal recognition of gender identity without imposing abusive requirements(Ghoshal & Knight, 2016). These international efforts are reinforced by resolutions such as the one issued by the Council of Europe in April 2015, which urges governments to adopt gender recognition procedures based on self-determination. Because the importance of legal recognition of gender identity transcends mere formality; it is a fundamental right to live a life of dignity, free from marginalisation. Enabling transgender people to determine and register their gender is a crucial step towards personal autonomy and respect for individual identity (Ghoshal & Knight, 2016).

In the European Union, some countries are leading the way in advancing legislation in favour of the transgender community, such as Spain, which pioneered Legal Gender Recognition (LGR), which allows people to change their name and gender marker in official records. Although all EU Member States and the UK recognise some form of LGR, the requirements vary significantly ('Legal Gender Recognition in the EU', 2020). However, many countries are still lagging behind in addressing this important social issue. It is therefore vital to analyse public opinion and the socio-economic situation in the EU in order to understand the political and cultural contexts and to promote social developments effectively.

2. Analysis

Special Eurobarometer 493 on discrimination focuses on perceptions and attitudes towards discrimination on various grounds, including gender identity. In the following analysis we aim to provide a broad overview of the challenges and opportunities for equality and protection in the EU, especially for the transgender community.

For this purpose, a complete analysis will be carried out, where firstly, through a descriptive analysis, we will be able to get an idea of the socio-demographic profile of each European country that supports social change for transgender people.

For this purpose, a complete analysis will be carried out, where firstly, through a descriptive analysis, we will be able to get an idea of the socio-demographic profile of each European country that supports social change for transgender people. Then with a regression analysis the impact of different factors on support levels for the policy will be assessed to draw a clear conclusion of which factors are determining country differences. Lastly, several prediction models will be tested to finally choose the one we believe that could predict support levels for the policy best.

To answer the main problem of this study, we will use as a dependent variable, or target variable, the question that answers the level of support for transgender people to obtain official identity documents.

2.1. Data Pre-processing

In order to develop and optimise the analysis of the survey on discrimination against the trans minority, we first made an exhaustive selection of the most relevant and richest variables for our study. These were all socio-demographic and socio-cultural variables in order to get to know the individual and country level profile and, in relation to the topic in question, all variables whose content deals with discrimination or the rights of the group.

The next step was to categorize everything that is *refusal, spontaneous responses,* or *don't know,* is categorised as NA (missing value). After doing this, the missing values increased considerably. What to do with variables with such a high percentage of missing values? Are these variables significant or will they make our study biased? The solution to these questions was to determine a screening criterion, we decided to get rid of the variables with more than 10% NA. The variable ideology has 17% missing values. However, we consider it relevant for understanding people's attitudes towards the target variable.

Once we had the relevant and information-rich variables for our study, we decided to add variables to learn even more about the socio-economic situation of the countries we are observing. Therefore, we added the following variables:

- GDP per capita (constant 2015 US$)
- Political Stability and Absence of Violence/Terrorism: Estimate
- Unemployment, total (% of total labor force)
- Gini Index
- Life expectancy at birth, total (years)
- Government Expenditure on Education (% GDP)

And the last step in pre-processing data was to impute the NAs, in order to be able to work with the database clean of missing values. To do this, we did a multiple imputation with MICE, which operates under the assumption that missing values occur at random (Missing at Random), which implies that the probability of missing values is determined only by observable data, allowing predictions based on those observed values. When performing MICE, we could previously check which of the imputation methods was the most suitable for our data set, in our case we decided to opt for the Predictive Mean Matching (PMM) method. After imputation using PMM, the number of missing values decreased considerably for each variable, although after this, it was necessary to eliminate those missing values that remained.

3. Descriptive analysis: Cross country differences in support levels

3.1. Overview

To clarify and condense our variable target of assessing support for a Legal Gender Recognition (LGR) policy, we note that 60.24% support this measure, while 39.75% do not. To provide an initial view on the particular stance of each European country we have made a heatmap, which is provided in Figure 1 in the annex.

To deepen our understanding of attitudes towards Legal Gender Recognition (LGR) within the EU, we resorted to clustering analysis. This technique allowed us to organise member states into groups based on their similarities, not only in terms of their positions on the LGR, our dependent variable, but also considering a wide range of related variables. The results of this analysis provide valuable insights for addressing the issue of the LGR in a comprehensive manner.

Through the dendrogram we see that the countries that join at the lower parts have more similar profiles in terms of the variables mentioned, while those that join higher up differ more from each other. For example, BG (Bulgaria) and RO (Romania) form an early cluster, suggesting that they have similar characteristics and responses to the variables studied. In contrast, countries that join at a higher altitude, such as DE (Germany) and GB (UK), differ

more significantly, which is remarkably similar to the arrangement of countries found in the heatmap (Figure 1). Overall, we can see that there are distinct "families" of countries with similar levels of support and similar socio-economic and demographic characteristics with respect to the transgender population and the dependent variable of interest, which can be visualised in Figure 2 and in Figure 2, where the phylogenetic tree clustering has been depicted.

3.2. Individual-level analysis

Focusing on socio-demographic variables and those most pertinent to LGR support, we find that support for LGR policies is primarily associated with ideology and religion, which is highly logical, along with support for LGTBI rights, as depicted in Figure 4.

Regarding socio-demographic variables characterizing each country in relation to the dependent variable, we observe the following summarized trends: Support decreases with age, with a notable difference between younger and older generations. Regarding gender, there's no significant difference between groups, but women tend to support this policy more (with a mean of 0.63). Additionally, there's a positive association between socio-economic status and policy support, with support increasing as socio-economic status improves.

Ideology plays a crucial role, with individuals leaning towards left-wing ideologies (1-3) showing higher levels of LGR support compared to those with more conservative ideologies. Thus, the proportion of individuals supporting LGR policies tends to decrease as ideology becomes more conservative.

Religion is another influential factor, where Protestants and Catholics exhibit high levels of LGR support (69.9% and 57.2% respectively), while Orthodox Christians show lower support (42.6%). Atheists and agnostics demonstrate the highest support (73.2% and 72.5% respectively), followed by Buddhists, Sikhs, and Jews (61.1% to 75%). Muslims (both Shia and Sunni) show lower levels of support (45.8% and 55.9%).

Finally, the results suggest considerably lower support levels in rural areas compared to small/medium and large cities, with the highest support percentage coming from small and medium-sized cities.

3.3. Country-level analysis

At the country level, variables such as GDP, life expectancy, or political stability help us gain a more organized perspective toward consciously supporting the legalization of gender identification. A notable finding is that GDP shows a moderately strong positive correlation, suggesting that countries with higher GDP tend to have higher levels of support. Additionally, the data indicate a weak positive correlation (0.184) between stability and being in favor of political change. This implies that countries with greater political stability may slightly lean towards greater support for LGTBIQ policies. Similarly, unemployment shows a weak positive correlation, indicating a slight tendency towards higher levels of support for LGTBIQ policies in countries with high unemployment rates. Finally, the Gini coefficient suggests that countries with lower income inequality tend to have slightly higher levels of support for LGTBIQ policies.

Overall, the model highlights the importance of GDP, unemployment, and the Gini coefficient (and its logarithm) as predictors of support for LGTBIQ policies, while suggesting that stability may not be statistically significant in this regard (see Figure 5).

Furthermore, a naive Logistic regression is performed with the country-level variables added as sole predictors as well as other features created through feature engineering. Overall the model seems to find all variables significant except for GDP and Unemployment^2. Nonetheless this can change when new variables are added.

4. Multivariate analysis

In order to finally test which are the most relevant variables when it comes to supporting the policy we are studying, a logistic regression is performed with all the relevant and significant variables of our dataset with the following specification:

glm(formula = policy_id ~ age + gender + ideo + class + none_min + oncontent + age_ideology + suprights + schsexor + schtrans + phone + religion + GDP + I(GDP^2) + Stability + I(Stability^2) + lifexp, family = binomial(), data = df_final)

In this logistic regression model, several factors significantly influence the likelihood of the policy-related outcome (policy_id). Notably, older age, male gender, stronger ideological views, absence of minimum wage laws (none_min), lower oncontent scores, higher scores on suprights, schsexor, and schtrans, phone ownership, specific religious affiliations, higher GDP, lower stability, and higher life expectancy all correlate with varying changes in the log odds of having the policy-related outcome.

It stands out though that only class 5 (being part of the highest class) and religions 2, 7, 12 and 13 have a significant impact on the log odds of the policy being enacted.

4.1 Discussion: Differences by country

After analysing the impact of various individual-level and country-level variables on the support for a proposed LGBTQ policy, several key factors emerged:

- **Age**: Older individuals tend to be less supportive of the policy, but the effect size is relatively small.
- **Gender**: Males show lower support compared to females.
- **Ideology**: Stronger ideological views correlate with lower support for the policy, indicating a significant impact.
- **Class**: There's a trend where higher class individuals tend to support the policy more, with the Netherlands standing out.
- **None_min**: Those who don't consider themselves minorities are more supportive, consistent with the positive coefficient in the regression.
- **Oncontent**: Sharing content on discrimination doesn't have a clear relationship with support for the policy.
- **Suprights, schsexor, schtrans**: Lower support for LGBTQ rights, sex orientation education, and transgender education in schools aligns with lower support for the policy.
- **Having a phone:** Phone ownership doesn't seem to strongly correlate with support for the policy.
- **Religion**: Orthodox Christian-majority countries tend to oppose the policy more, while atheist and agnostic-majority countries are more supportive.
- **GDP, Stability, Life expectancy**: Countries with higher GDP, stability, and life expectancy tend to exhibit greater support for the policy, indicating a broader trend towards progressiveness in developed nations.

Having said this, we would determine that what mainly explains the differences in support between countries are country-level factors such as GDP, Stability and Life expectancy as well as individual-level indicators related somehow to ideology such as religion, ideology itself and support for equality of rights with the LGTB community and for the inclusion of sexual orientation with diversity terms and transgender information lessons at schools (Figs. 6-12).

5. Prediction

Today, advocating for gender identity rights is a critical issue. One significant aspect of this advocacy is the ability of transgender individuals to change civil documents to align with their gender identity. To understand and predict the levels of support for such rights in various countries, machine learning techniques are employed. This task involves the development and evaluation of predictive models using factors identified in prior research. In this context, three predictive models - Logit, Random Forest, and LDA (Linear Discriminant Analysis) - have been developed and will be assessed for their efficacy in predicting support for transgender individuals' rights to change civil documents to align with their gender identity across different countries.

Our interpretation for the models is the following:

- The logit model's accuracy is around 75%. This means that out of all the predictions made by the model, around 75% of them were correct. Overall, this is a good model.
- The accuracy of the Random Forestmodel is approximately 77%. This indicates that about 77% of the predictions made by the model are correct. Sensitivity measures the proportion of actual positive cases that were correctly identified by the model. In this case, it's approximately 84%. Specificity measures the proportion of actual negative cases that were correctly identified by the model. In this case, it's approximately 67%. This seems like a better model than the previous.
- Finally, the LDA model seems like a better fit for our data. The accuracy of this model is around 85%, with higher specificity and sensitivity levels. Therefore, the LDA seems like the best model for predicting the levels of support towards the mentioned gender identity policy.

6. Conclusion

In summary, the comprehensive analysis of support for gender identity recognition policies reveals a number of important findings at both the individual and country level. At the individual level, factors such as age, gender, ideology and religion show a significant influence on support for these policies, with a trend towards majority support among young people, women, individuals with progressive ideologies and those with no religious affiliation. At the country level, variables such as GDP, political stability and life expectancy show positive correlations with support for these policies, suggesting that countries with better socio-economic indicators tend to be more progressive in terms of gender identity rights.

In addition, by using machine learning techniques, predictive models have been developed that demonstrate the ability to forecast levels of support for gender identity policies in different countries. Although all three models - Logit, Random Forest and LDA - show some predictive efficiency, the LDA model stands out as the most accurate, with an accuracy rate of 85% and high levels of sensitivity and specificity.

Overall, our analysis successfully identified patterns of behavior, both in the individual and country levels, for understanding and predicting the levels of support towards the gender identiy policy of interest.

References

Ghoshal, N., Knight, K., & HUMAN RIGHTS WATCH. (2016). Rights in Transition: Making Legal Recognition for Transgender People a Global Priority. In *World report 2016: Events of 2015* (1st ed., pp. 21–32). Bristol University Press. https://doi.org/10.2307/j.ctvndv9bj.5

Legal gender recognition in the EU. (2020). *Publications Office of the European Union*.

United Nations Resolutions on sexual orientation, gender identity and sex characteristics. (n.d.). OHCHR. Retrieved 20 March 2024, from https://www.ohchr.org/en/sexual-orientation-and-gender-identity/united-nations-resolutions-sexual-orientation-gender-identity-and-sex-characteristics

Van Der Vleuten, A. (2020). Contestations of transgender rights and/in the strasbourg court. *Politics and Governance*, *8*(3), 278–289. https://doi.org/10.17645/pag.v8i3.2876

Appendices

Figure 1: Proportions of Support for the Dependent Variable, by country

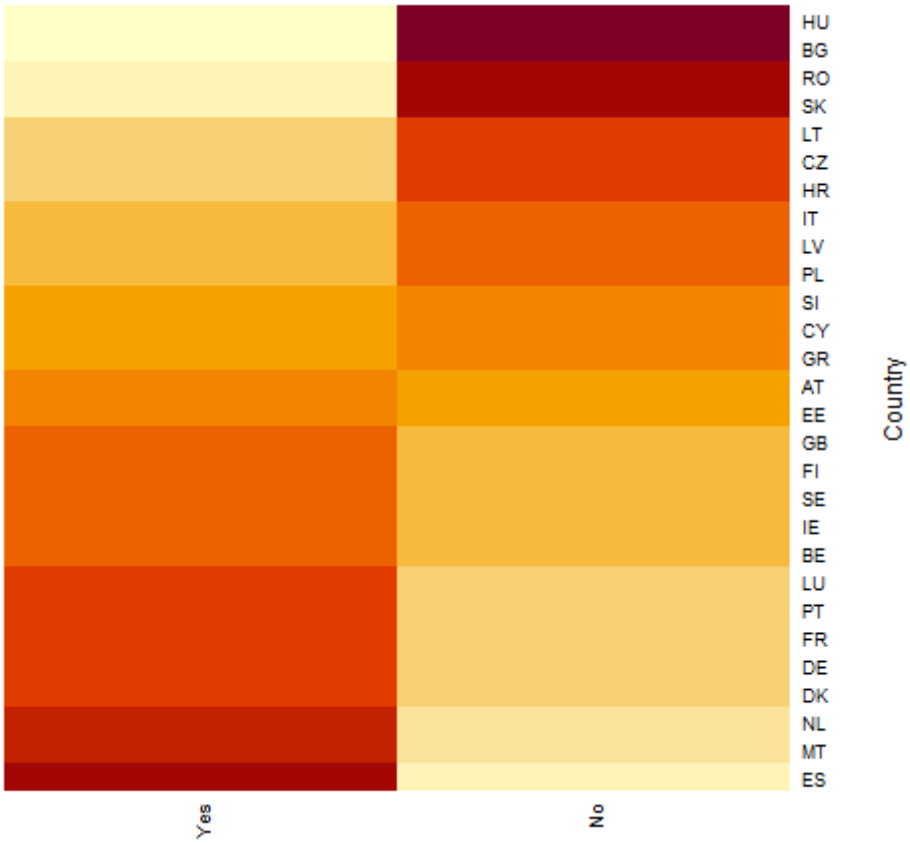

Figure 2: Dendrogram

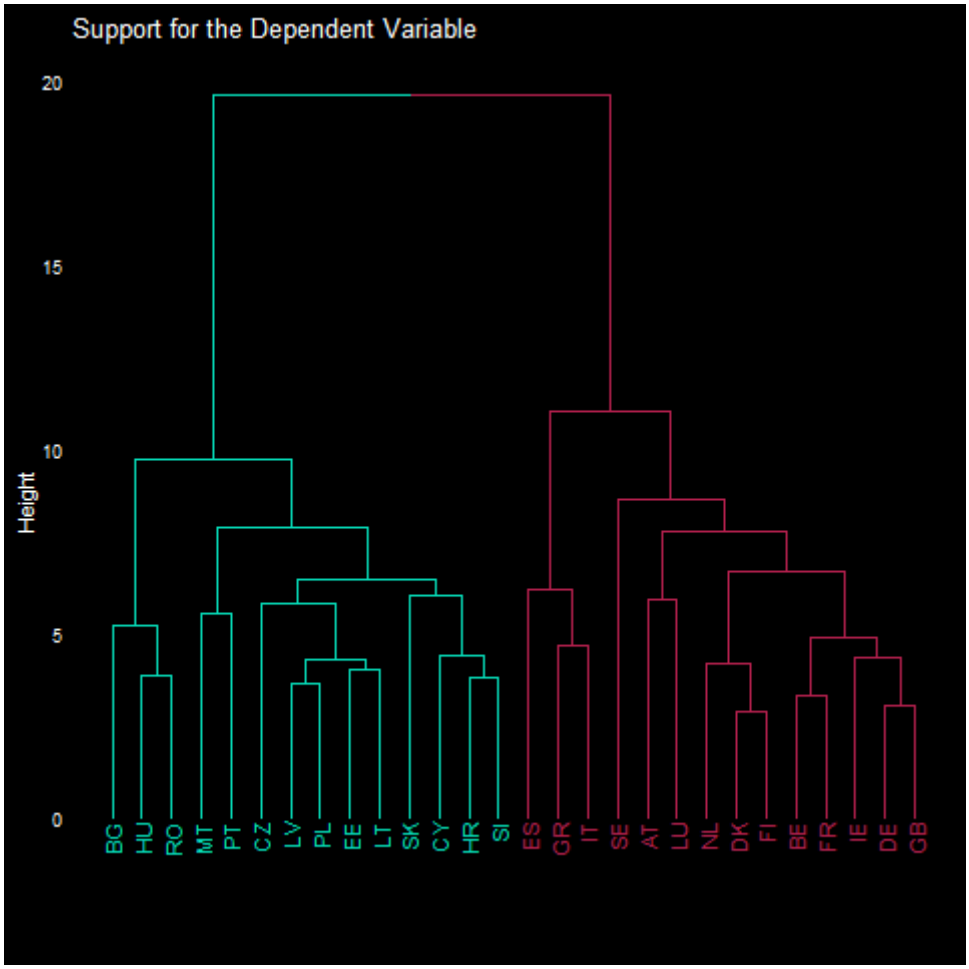

Figure 3: Phylogenic tree

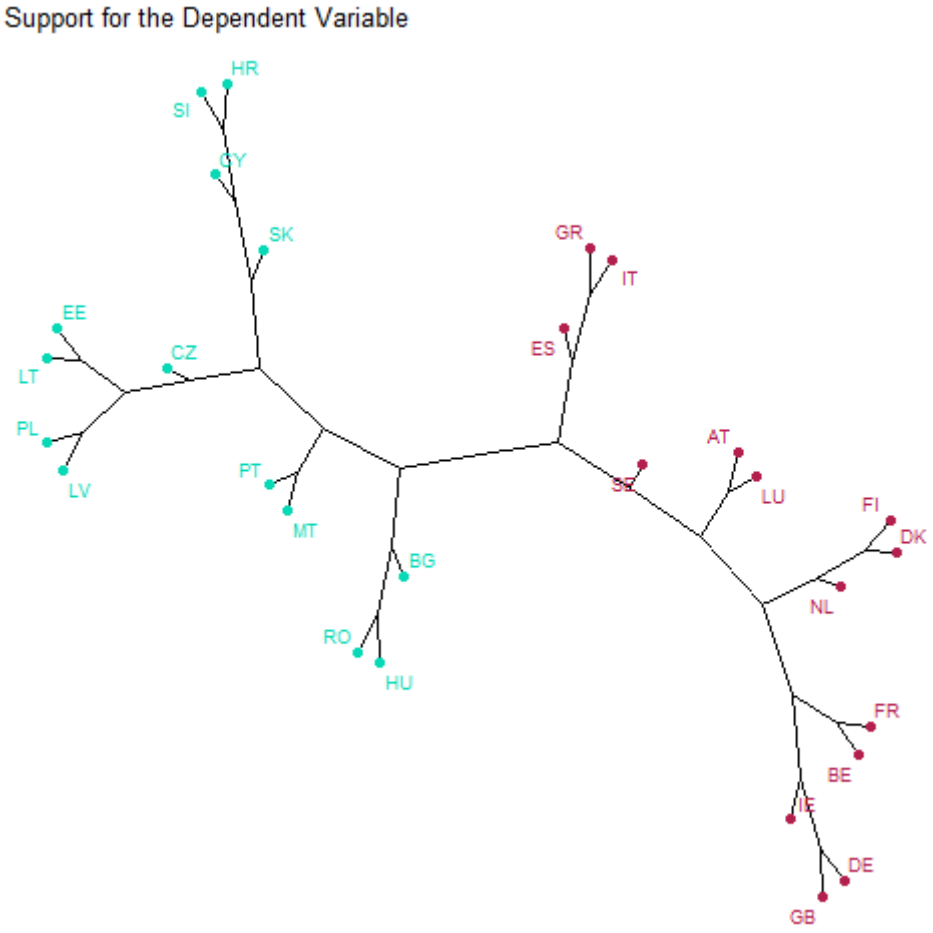

Figure 4: Correlation Sociodemographic Variables

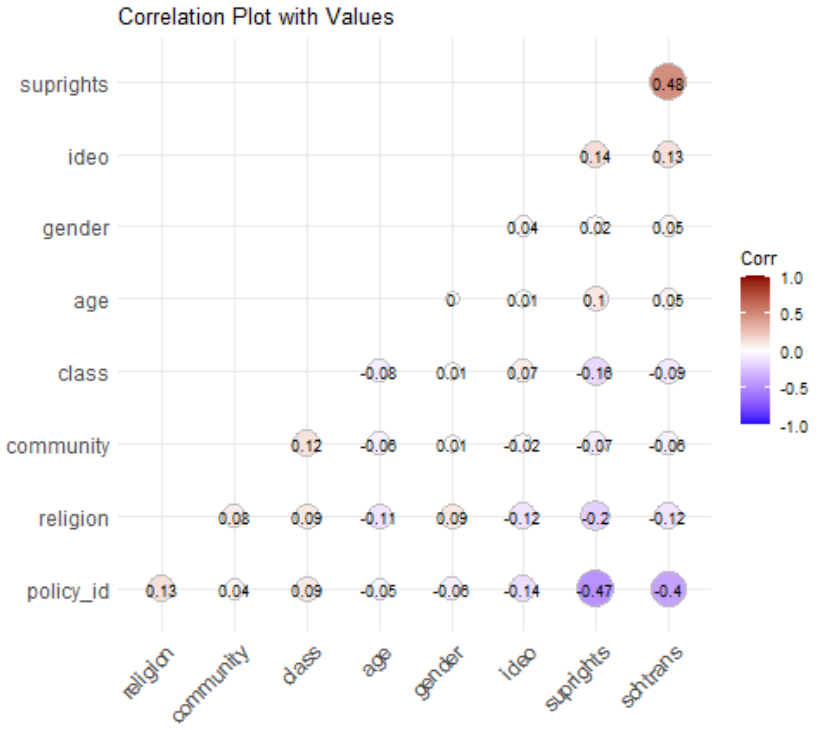

Figure 5 Correlation Country Level

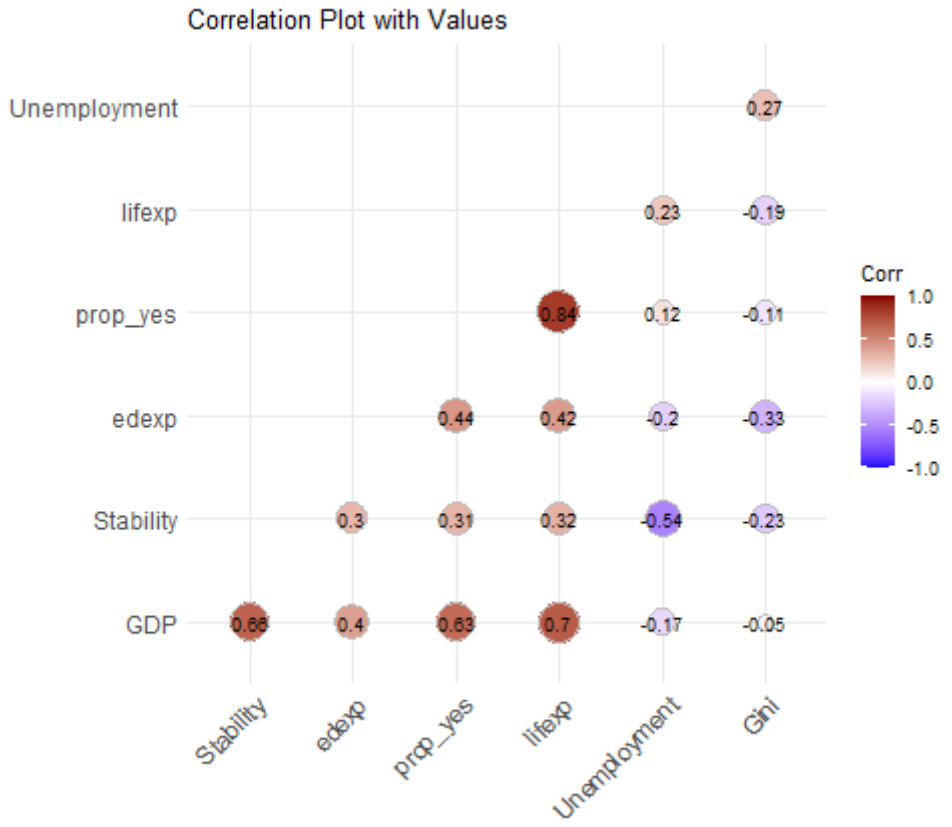

Figure 6: GDP by Country

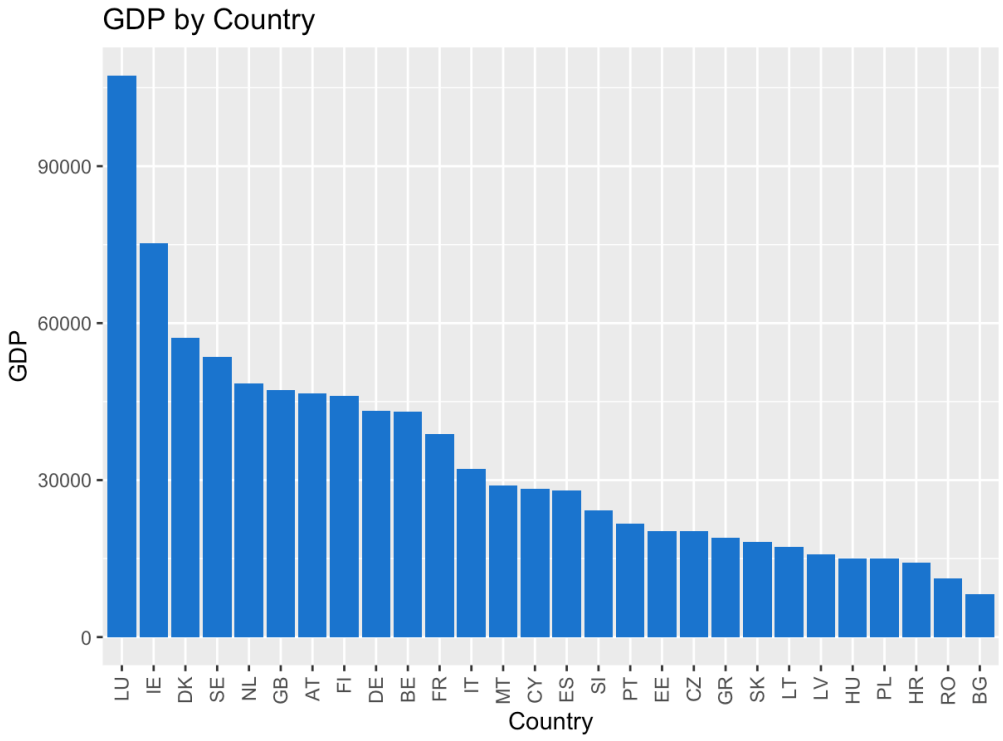

Figure 7: Idology by country

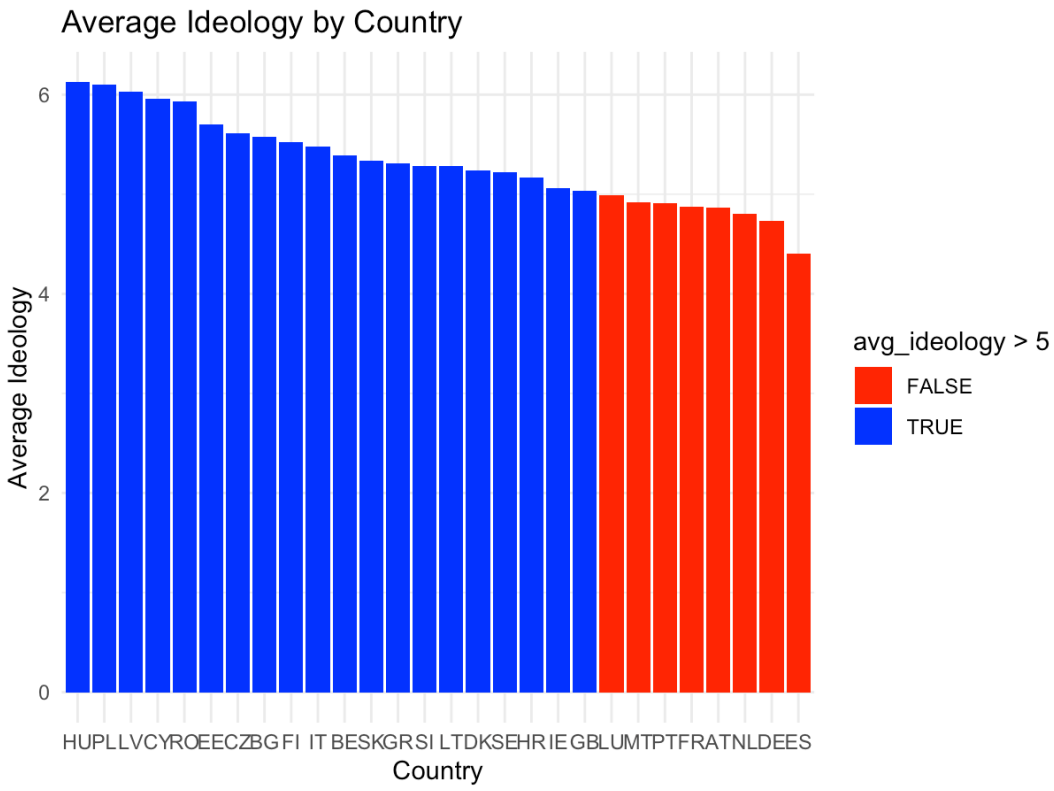

Figure 8: Life expectancy by country

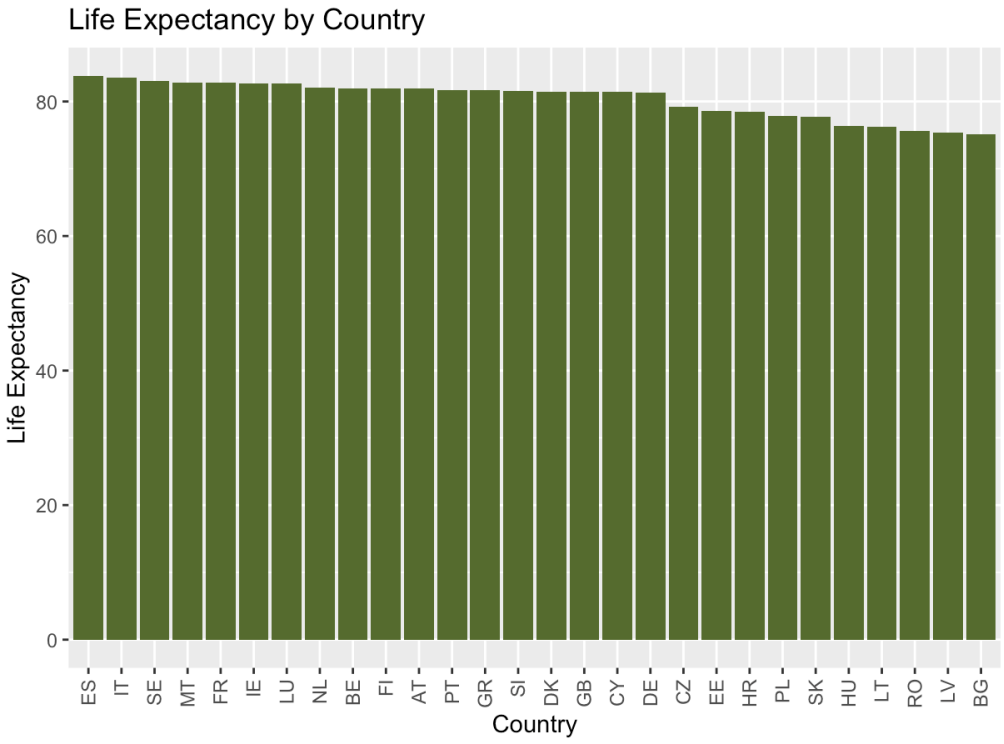

Figure 9: Religion by country

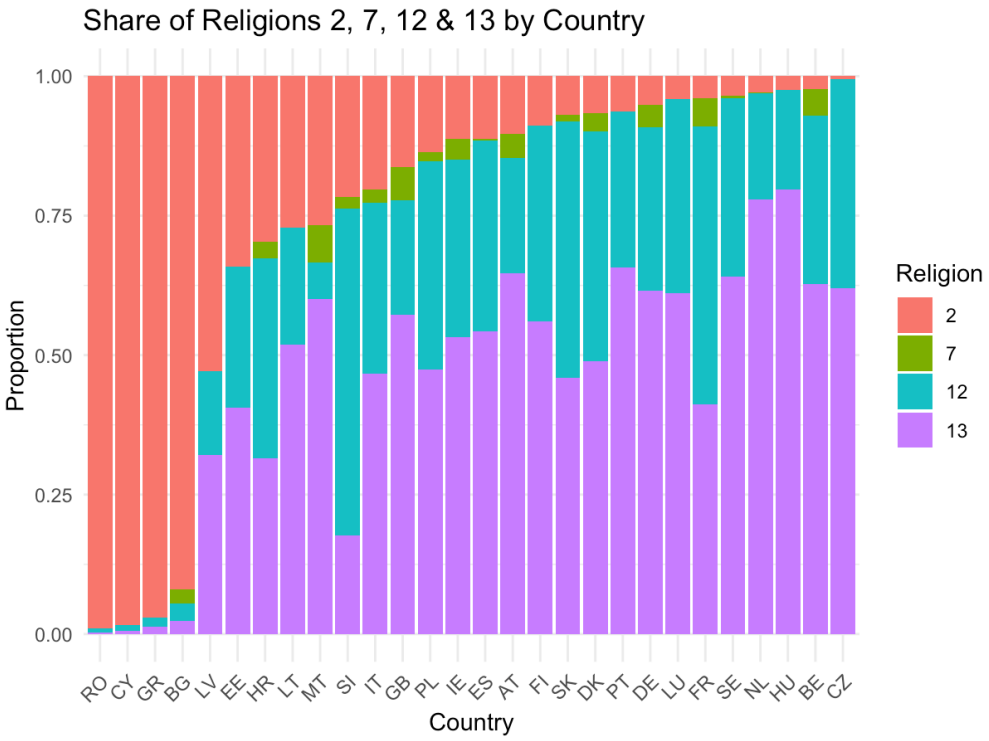

Figure 10: Satibility by country

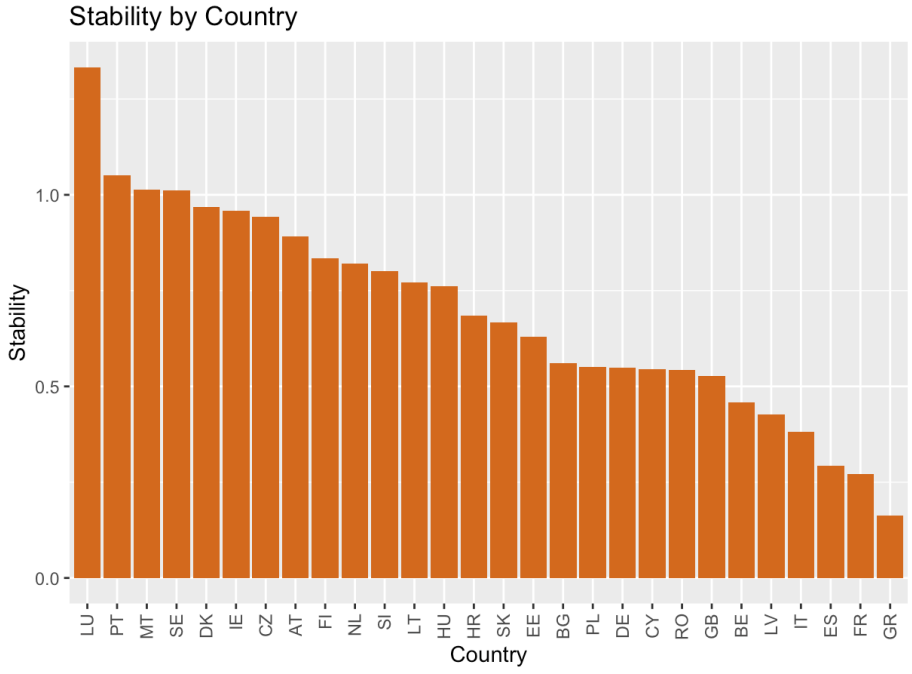

Figure 11: Relevant variables by country

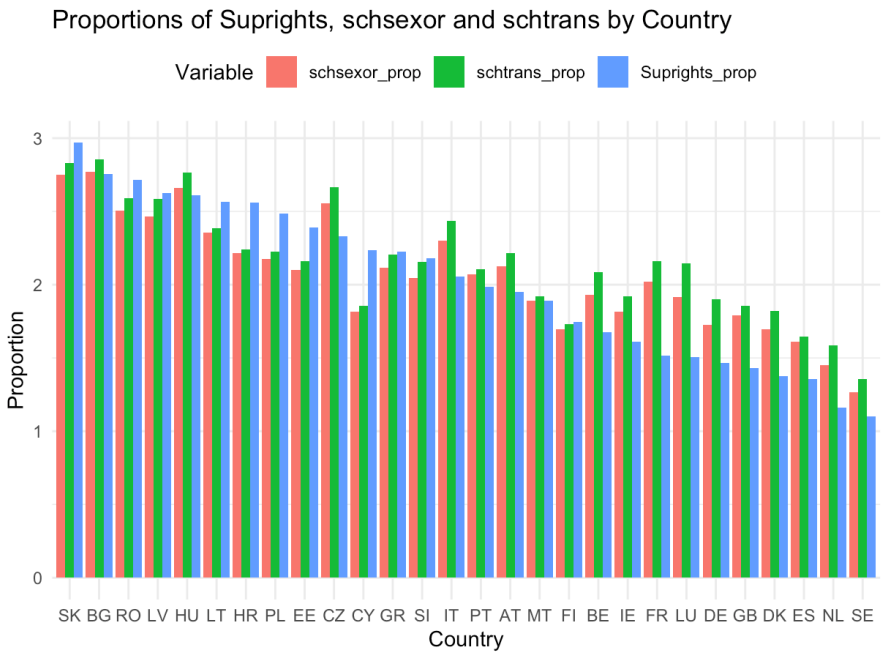

Figure 12: Proportion of support by country